



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : [http://oatao.univ-toulouse.fr/Eprints ID : 15363](http://oatao.univ-toulouse.fr/Eprints/ID/15363)

The contribution was presented at SUM 2015:
<http://www.fernuni-hagen.de/wbs/sum2015/>

To cite this version : Bisquert, Pierre and Croitoru, Madalina and Dupin De Saint Cyr - Bannay, Florence *Towards a dual process cognitive model for argument evaluation*. (2015) In: 9th International Conference on Scalable Uncertainty Management (SUM 2015), 16 September 2015 - 18 September 2015 (Quebec, Canada).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Towards a dual process cognitive model for argument evaluation*

Pierre Bisquert¹ Madalina Croitoru² Florence Dupin de Saint-Cyr³

¹ INRA, France pierre.bisquert@supagro.inra.fr

² Univ. Montpellier, France croitoru@lirmm.fr

³ IRT, France florence.bannay@irit.fr

June 30, 2015

Abstract

In this paper we are interested in the computational and formal analysis of the persuasive impact that an argument can produce on a human agent. We propose a dual process cognitive computational model based on the highly influential work of Kahneman and investigate its reasoning mechanisms in the context of argument evaluation. This formal model is a first attempt to take a greater account of human reasoning and is a first step to a better understanding of persuasion processes as well as human argumentative strategies, which is crucial in collective decision making domain.

Keywords: Cognitive computational models, dual process reasoning, persuasion, argument.

1 Introduction

Gaining more and more attention, persuasion is a crucial aspect of human interaction and is closely linked to social groups creation and dynamics [30, 33]. With the recent rise of computer science technology, the study of persuasion began to transcend its original fields (including psychology, rhetoric and political sciences) and to take lasting root in the artificial intelligence (AI) domain.

In the AI domain, two predominant trends may be identified: *interactive technologies for human behavior* and *dialogue protocols for persuasion*. The former trend aims at producing systems able to persuade humans to change their behavior for another one considered better [21]. It has often been used in the context of health-care [19], environment [5] or education [12]. Such an approach, by definition, is human-machine oriented. The latter trend, derived from logic and philosophy authors such as Hamblin [13], Perelman [22] or Walton [31], aims at creating normative dialogue protocols ensuring rational interactions between agents [20, 1, 24]. The proposed protocols regulate the persuasion processes engaged between agents such that conflicts are resolved in a fair manner. These approaches are often machine-machine oriented and prescriptive.

In this paper we are interested in the computational and formal analysis of the persuasive interactions that occur between humans. Since humans are known to be subject to reasoning biases, we are interested in the link between persuasion and cognitive biases. The importance of this subject has, in particular, been highlighted in the field of law in the context of a court [8] or psychology [15]. This formalisation is a first step towards a better understanding of human persuasion strategies and may help to detect and notify cognitive biases, e.g. in protocols handling collective decision making.

Several works in psychology analyze cognitive biases with the help of *dual process theory* [29, 9, 11, 2, 10, 26], where reasoning may be achieved thanks to two different processes, one being heuristic,

superficial and fast, and the other being scrupulous, thorough and slow. Indeed, according to Kahneman [29], the first system (called S1) deals with quick and instinctive thoughts and is based on associations such as cause-effect, resemblance, valence, etc. The second system (called S2) is used as little as possible and is a slow and conscious process that deals with what we commonly call reason. Cognitive biases arise mostly when the superficial reasoning is used. In their seminal article [29], Tversky and Kahneman explain how supposedly “rational” judgments are based on data with limited validity and processed according to heuristic rules. They illustrate their thesis with a number of biases empirically demonstrated (such as the illusion of validity, retrievability of instances, anchoring, framing, etc.). This diptych has been popularized in many domains including persuasion [23, 6]. In the Elaboration Likelihood Model [23], two *routes* might be used to persuade someone: the *central route*, which calls for a careful examination of the received message, and the *peripheral route*, using simple cues to evaluate the position advocated by an orator. While works such as [29] and [23] coincide in spirit, our aim is to unify them into a formal framework with four cognitive profiles for evaluating an argument such that a more engaged agent will use a deeper reasoning (S2) while a quiescent agent will only use associations (S1).

After defining a new cognitive model and two reasoning processes based on [29] as well as [23] in Section 2, we present how an argument might be evaluated and its effect on the agent’s mind in Section 3. Finally, some properties are shown in Section 4.

2 Towards a Computational Model of Cognitive Evaluation

2.1 Cognitive Model

In this paper, our aim is to define a computational cognitive model of the evaluation of an argument. Based on Kahneman’s theory, we propose to define an agent cognitive model as two components: AT (an association table linking a formula to an ordered set of formulae and to a flag encoding an appreciation) and KB (a logical knowledge base) in order to encode S1 and S2 respectively.¹ Formally, we consider a propositional language and we denote by \mathcal{L} the set of well formed formulae of this language given the usual connectives $\wedge, \vee, \rightarrow, \neg$ and the constants \perp and \top . The set of symbols in the language is denoted by \mathcal{V} . \vdash denotes classical inference. The fact that a symbol s appears in a formula φ is denoted by $s \in \varphi$. We also consider a propositional language, denoted \mathcal{L}_G , based on a set of symbols \mathcal{V}_G distinct from \mathcal{V} ($\mathcal{V}_G \cap \mathcal{V} = \emptyset$). Formulae of \mathcal{L}_G are called *generic formulae*.

Definition 1 (Association table) *An agent’s association table AT is a set of triples of the form $(\varphi, (S, \succ_S), f)$ where:*

- $\varphi \in \mathcal{L}$ is a well formed formula representing a piece of knowledge,
- $S \subseteq \mathcal{L}$ is a set of well formed formulae associated to φ endowed with a total strict order $\succ_S \subseteq S \times S$, the pair (S, \succ_S) is called a *stack* (when there is no ambiguity, the total order will be omitted),
- $f \in \{\oplus, \ominus, \odot\}$ is a flag stating that φ is respectively accepted, rejected or not specified (also called empty flag) in the association table.

The set of all well formed formulae in the association table is denoted by L_{AT} , i.e., $L_{AT} = \bigcup_{(\varphi, (S, \succ_S), f) \in AT} \{\varphi\}$. Given a formula $\varphi \in L_{AT}$, the stack S associated with φ in AT will be denoted by $AT(\varphi)$, the i^{th} element of S is denoted $AT(\varphi, i)$, and the top element of this stack is denoted $\text{Top}(\varphi)$ ($\text{Top}(\varphi) = AT(\varphi, 1)$). Formally, $\text{Top}(\varphi) = \varphi_0$ s.t. $\forall \varphi' \neq \varphi_0 \in AT(\varphi), \varphi_0 \succ_S \varphi'$. The flag f associated to φ is denoted by $\text{flag}(\varphi)$. If f is a flag then $-f$ is a flag such that $-\oplus = \ominus, -\ominus = \oplus$ and $-\odot = \odot$. Note that AT is implicit in the definitions of Top and flag .

A knowledge base contains Strict and Defeasible Beliefs, Appreciations (i.e. associations of formulae to flags) and a set of Appreciation Rules² called *a-rules* as described below.

Definition 2 (Knowledge base) *A knowledge base KB built on \mathcal{L} and \mathcal{L}_G is a quadruplet $KB = (F, \Delta, A, R)$ s.t. $F \subseteq \mathcal{L}$ is a set of formulae, Δ is a set of default rules, A is a set of appreciations and R is a set of a-rules, where*

¹Note that S1 and S2 are linked as we will see in (1) of Definition 3.

²Inspired from the Desire-Generation rules (of Rahwan and Amgoud [25]).

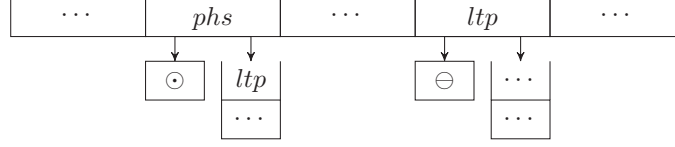


Figure 1: Partial representation of the associative table.

- A default rule is denoted $a \rightsquigarrow b$ with $(a, b) \in \mathcal{L} \times \mathcal{L}$ with the intended meaning “if a is true then generally b holds”.
- An appreciation is a pair $(\varphi, f) \in \mathcal{L} \times \{\oplus, \ominus, \odot\}$ meaning that φ is associated to the flag f .
- An a-rule has the form $(E_K, E_A) \Rightarrow (\psi, f)$ where $E_K \subseteq \mathcal{L}_G \times \mathcal{L}_G$ is a set of pairs of generic formulae (called generic default rules), $E_A \subseteq \mathcal{L}_G \times \{\oplus, \ominus, \odot\}$ is a set of generic appreciations, $\psi \in \mathcal{L}_G$ is a generic formula and $f \in \{\oplus, \ominus\}$ is a flag. This kind of rule has the intended meaning “if all the default rules E_K apply in a given context and if all the appreciations E_A hold then generally the new appreciation (ψ, f) is valid”.

The use of default rules has two main interests. First, it simplifies the writing: it allows us to express a rule without mentioning every exception to it. Second, it allows us to reason with incomplete descriptions of the world: if nothing is known about the exceptional character of the situation, it is assumed to be normal, and reasoning can be completed.

Definition 3 (Cognitive model) A cognitive model is a tuple $\kappa = (KB, AT, \lambda, i)$:

- $KB = (F, \Delta, A, R)$ is a knowledge base,
- AT is an association table such that: $\forall \varphi, \varphi' \in \mathcal{L}, \forall f \in \{\oplus, \ominus, \odot\}$,

$$\left. \begin{array}{l} - \text{ if } \varphi \in F \text{ then } \forall s, s' \in \varphi, s \in AT(s'), \\ - \text{ if } \varphi \rightsquigarrow \varphi' \in \Delta \text{ then } \varphi' \in AT(\varphi), \\ - \text{ if } (\varphi, f) \in A \text{ then } \text{flag}(\varphi) = f, \end{array} \right\} \quad (1)$$
- $\lambda \in \mathbb{N}$ is an integer value representing the threshold above which the agent feels to be enough aware about the topic of a formula to be able to reason rationally,
- $i : \mathcal{L} \rightarrow \{0, 1, 2\}$ is a three value marker that gives the interest level of the agent relatively to a formula.

In other words, (1) expresses the link between KB and AT , more precisely, every pair of symbols belonging to a given formula in F , and every pair of formulae in Δ linked by a default rule, are associated in AT and the flags in AT comply with A . In case of ambiguity about the current cognitive model, the symbols AT , Top , flag will be indexed by the cognitive model κ they refer to.

Example 1 We illustrate here the question of performing the separation of durum wheat cereal (or other plants in the field such as peas) after the harvest that was done within an ANR DUR-DUR³ meeting. As our keen internship student was performing his literature review, he quickly learned that post harvest separation (phs) is efficient (eff), which implies a process that is not expensive ($\neg exp$). His KB contains formulae such as $phs \rightsquigarrow eff$ and $eff \rightsquigarrow \neg exp$. However, during a coffee break, he heard a colleague working on post harvest separation with optical harvest devices (opt) and learned that these instruments are generally very long to produce (ltp): $phs \wedge opt \rightsquigarrow ltp$. He is certain that long production is not efficient: $ltp \rightarrow \neg eff$. While he still does not know whether to accept or reject the post harvest separation, the first thing he now associates post harvest with is the long time to produce, something he disapproves of. This is represented by the flag \ominus in AT (see Fig. 1) and by the appreciation (ltp, \ominus) in KB .

2.2 System 1 and System 2 Reasoning

Let us see how to use this representation framework in order to reason. In this paper, we call *reasoning* the process of evaluating the acceptability of a formula $\varphi \in \mathcal{L}$, i.e., mapping φ to a flag in $\{\oplus, \ominus, \odot\}$. The reasoning is not the same in S1 and S2. In S1, reasoning is based on the association table AT while in S2

³French funded project aiming at improving durum wheat sustainability (<http://www.agence-nationale-recherche.fr/?Project=ANR-13-ALID-0002>).

it is based on an inference principle. We propose to encode S1-reasoning as follows: if the current formula has a non-empty flag, then this flag is returned; else, if the current concept has an empty flag, the concepts of the stack associated to the current concept are evaluated recursively, in an order relative to their position in the stack and the number of iterations.

We first define a *reflection path* R_φ associated to a concept φ thanks to a sequence D_φ of iterations from the initial formula φ . This sequence contains the successive depths d_i in the stacks corresponding to formulae with an empty flag that are necessary to follow in order to find a formula with a non-empty flag. The reflection path jumps recursively from a formula φ_i to a formula φ_{i+1} if φ_{i+1} appears in the stack of φ_i at the depth d_i (each depth d_i in the sequence should not exceed the total depth of each stack $AT(\varphi_i)$). Note that many reflection paths can be built from a formula φ ; this is why we will select the cheapest one in terms of cognitive effort.

Definition 4 (Reflection path) A reflection path $R_\varphi = (\varphi_1, \dots, \varphi_n)$ from φ is a sequence of $n \geq 1$ formulae corresponding to a sequence (d_1, \dots, d_{n-1}) of $n - 1$ integers such that $\varphi_1 = \varphi$ and recursively

$$\forall 1 \leq i < n, \quad \varphi_{i+1} = AT(\varphi_i, d_i), \quad \text{with } d_i \leq |AT(\varphi_i)| \text{ and } \text{flag}(\varphi_i) = \odot.$$

We denote $\text{flag}(R_\varphi)$ the flag associated to the last concept reached by the sequence R_φ , hence, $\text{flag}(R_\varphi) = \begin{cases} \text{flag}(\varphi_n) \neq \odot & \text{if } n \text{ is finite,} \\ \odot & \text{otherwise.} \end{cases}$

The cognitive weight associated to a reflection path $R_\varphi = (\varphi_1, \dots, \varphi_n)$ associated to the integers $(d_i)_{1 \leq i < n}$ is $\text{weight}(D) = \sum_{i=1}^{n-1} d_i + n$.

The *cognitive weight* associated to a sequence allows to take into account both the depth in the stack and the number of iterations. The more deep and long is the sequence, the more it requires an effort to the agent. S1-reasoning will amount to find and follow reflection paths of *minimal cognitive weight* until a non-empty flag is reached. Hence, S1-reasoning consists in finding a non-empty flag to associate to a concept while minimizing⁴ the cognitive effort.

Definition 5 (S1-reasoning) Given a cognitive model $\kappa = (KB, AT, \lambda, i)$, We call S1-entailment, denoted by \vdash_1 , the inference obtained by following a reflection path: $\varphi \vdash_1 \psi$ iff $\psi \in R_\varphi$ and R_φ is finite.

We define S1-reasoning⁵, about a formula φ , denoted $\text{eval}_1(\varphi, \kappa)$, as $\text{eval}_1(\varphi) = \text{flag}(R_\varphi)$ where R_φ is a reflection path from φ s.t. there is no reflection path R'_φ from φ with $\text{weight}(R'_\varphi) < \text{weight}(R_\varphi)$.

Example 2 Given the association table shown in Fig. 1, the result of $\text{eval}_1(phs)$ is \odot . Indeed, since the formula phs has the flag \odot , the S1-reasoning gets the top formula of the stack associated to phs , which is ltp ; the reflection path is $R_{phs} = (phs, ltp)$ and its associated sequence is (1) . The flag of ltp being different than \odot , it is the result of the evaluation.

Concerning S2, the study of the best rational model among all the proposals done in the AI literature is out of the scope of the paper. We propose to use, for the sake of illustration, the idea of defeasible approach of [3], called “contextual entailment” which is an extension of the “preferential entailment” [17]. Preferential entailment is an inference relation satisfying “desirable” postulates (listed in Section 4).

The set of conclusions that one can obtain by using a “preferential entailment” is usually regarded as the minimal set of conclusions that any reasonable non-monotonic consequence relation for default reasoning should generate. Moreover, it correctly addresses the specificity problem: results issued from sub-classes override those obtained from super-classes [28]. Unfortunately, in spite of these two advantages, “preferential entailment” is too cautious and suffers from the so-called irrelevance problem: from a rule “generally,

⁴Note that we could also have given more weight to the depth in the stack than to iteration or conversely, hence transform the equation into $\text{weight}(D) = \alpha \cdot \sum_{i=1}^n d_i + \beta \cdot n$ with a “smart” tuning of the ratio between α and β (this tuning should be based on psychological experiments).

⁵In practice, a constructive method to obtain R_φ could be an adaptation of Dijkstra algorithm on a graph where the vertices are partial reflection paths. An arc would link a vertex to another vertex if it corresponds to an extension of the path of one iteration (hence there would be as many arcs starting from a given vertex as the stack corresponding to this vertex is deep), namely there would be an arc between (φ_1, φ_2) and $(\varphi_1, \varphi_2, \varphi_3)$. The algorithm should start from the vertex corresponding to the empty path (*i.e.* it corresponds to the initial concept φ) and find a shortest path to a vertex with a non-empty flag. The length of a path would be the weight of the reflection path R_φ contained in the last vertex of the path.

if a then b ", it is not possible to deduce that b follows from $a \wedge d$ even if d is irrelevant to a and b . A typical example of irrelevance problem is that from "generally, birds fly" it is not possible to deduce that "red birds fly".

The approach proposed in [3] has shown to be an extension of "preferential entailment" which corrects this problem. This is why we choose to build S2 on the same idea. This is based on the identification of default rules having exceptions in a given context:

Definition 6 ([3]) Let c be a consistent formula considered as the current context, let Δ be a set of default rules. A default rule $a \rightsquigarrow b \in \Delta$ has an exception with c if and only if one of the two conditions holds:

1. $a \wedge c \wedge b$ is inconsistent,
2. $\exists \varphi \in \mathcal{L}, c \vdash \varphi$ and $a \wedge \varphi \vdash_{\Delta} \neg b$,

where \vdash_{Δ} is the inference relation defined by the closure of the preference entailment relation \vdash_{\sim} over the set obtained by interpreting each default $a \rightsquigarrow b \in \Delta$ as $a \vdash b$.

Definition 7 (S2-entailment) Given a knowledge base $KB = (F, \Delta, A, R)$, S2-entailment, denoted \vdash_2 , is defined by $\forall \varphi, \varphi' \in \mathcal{L}, \varphi \vdash_2 \varphi'$ iff $F_{\varphi} \cup \{\varphi\} \not\vdash \perp$ and $F_{\varphi} \cup \{\varphi\} \vdash \varphi'$, where $F_{\varphi} = F \cup \{a \rightarrow b \mid a \rightsquigarrow b \in \Delta \text{ has no exception with } \varphi\}$.

Example 3 The student's KB is s.t. $\Delta = \{phs \rightsquigarrow eff, eff \rightsquigarrow \neg exp, phs \wedge opt \rightsquigarrow ltp\}$ and $F = \{ltp \rightarrow \neg eff\}$. It holds that $phs \vdash_2 \neg exp$ (using Cautious monotony on $phs \rightsquigarrow eff$ and $eff \rightsquigarrow \neg exp$ and Cut on $phs \wedge eff \rightsquigarrow \neg exp$ and $phs \rightsquigarrow eff$ and due to the fact that Contextual entailment generalizes Preferential entailment, see Proposition 1).

Note that we are not yet in position to define S2-reasoning, which could evaluate the flag of a formula φ given a cognitive model κ . In order to do so we should define an aggregation function that combines all the possible flags that could be obtained for φ given the available beliefs, appreciations and a-rules. However, we have enough material to define the evaluation of one argument as shown in the next section.

3 Argument Evaluation

3.1 Argument and Profiles

We first give a (restrictive) definition of an argument, since we only consider arguments in favor of appreciations and not in favor of beliefs as it is the case in, for instance, [1].

Definition 8 (Argument) An argument is a tuple $(s, h, w, (c, f))$ where s is a formula (the speaker enunciating the argument), h is a pair (K_h, A_h) with a set of default rules K_h and a set of appreciations $A_h \subseteq \mathcal{L} \times \{\oplus, \ominus, \odot\}$ (the premise of the argument), w is an a-rule (the warrant), c is a formula (the conclusion) and $f \in \{\oplus, \ominus\}$ is a flag stating that the argument conclusion should be accepted or rejected.

This definition is syntactic. Hence, quadruplets containing premises not linked with the conclusion may comply with our definition. It is up to the listener to declare if the argument is valid semantically. This is the aim of this section. In the ELM model [23], the determination of the "route" for persuasion is made thanks to two main factors: the *interest* in processing the message and the *ability* (wrt knowledge and cognitive availability) to process it. In our model, the **interest** is given by the function i (see Def. 3). An agent may be not interested by a formula φ ($i(\varphi) = 0$), interested ($i(\varphi) = 1$) or "fanatic" ($i(\varphi) = 2$). The **knowledge** is represented by the size of the stack related to φ in AT . This size is compared to the agent's threshold λ (see Def. 3) in order to link the quantity of information the agent has to his feeling about being sufficiently aware on φ .

We use these factors for distinguishing several profiles of agents (note that we leave the cognitive availability for future work). In order to make a clear-cut categorisation of the possible engagements and to comply with the notions used in the ELM model, we define four levels of engagement: unconcerned, enthusiastic, quiescent or engaged with increasing involved level of cognition (see Definitions 11-14). Such profiles represent typical (and extreme) dispositions wrt the evaluation of an argument which goes beyond the classical idea to propose credulous and sceptical attitudes (see e.g. [1]).

Definition 9 (Profile) The profile of an agent is a function that maps a formula $\varphi \in \mathcal{L}$ and a cognitive model $\kappa = (KB, AT, \lambda, i)$ to an element of $\{unc, ent, qui, eng\}$:

$$\text{profile}(\varphi, \kappa) = \begin{cases} unc & \text{if } i(\varphi) = 0 \\ qui & \text{if } i(\varphi) = 1 \text{ and } |AT(\varphi)| < \lambda \\ eng & \text{if } i(\varphi) = 1 \text{ and } |AT(\varphi)| \geq \lambda \\ ent & \text{if } i(\varphi) = 2 \end{cases}$$

The following postulate expresses that if an agent is **enthusiastic** about a formula φ , then she has an opinion about φ .

Postulate 1 $\text{profile}(\varphi, \kappa) = ent$ implies $\text{flag}_\kappa(\varphi) \neq \odot$.

The next section details the value of the function evalarg defined below.

Definition 10 (Evaluation of an argument) Given a cognitive model $\kappa = (KB, AT, \lambda, i)$, an argument $a = (s, h, w, (c, f))$ and a profile $p = \text{profile}(c, \kappa)$, let evalarg be a function that maps a and p to an evaluation of the argument in $\{\oplus, \ominus, \odot\}$, denoted as $\text{evalarg}(a, p)$.

3.2 Argument Evaluation According to Profiles

In this section, we define formally how the evaluation is done with respect to the four profiles.

Unconcerned As its name implies, the unconcerned profile represents the fact that no interest is given by the agent in the received argument. Hence, an unconcerned agent will not bother trying to evaluate this argument and will just discard it.

Definition 11 (Unconcerned evaluation) Given an argument $a = (s, h, w, (c, f))$, the evaluation of a by an unconcerned agent unc is never done.

Enthusiastic The enthusiastic profile represents the fact that an agent is already convinced. As such, she does not feel the need to evaluate rationally the argument and will just check if the flag of the argument's conclusion correspond to the flag in her AT .

Definition 12 (Enthusiastic evaluation) Given an argument $a = (s, h, w, (c, f))$, the evaluation of a by an enthusiastic agent $\text{evalarg}(a, ent) = \oplus$ iff $\text{eval}_1(c) = f$ else $\text{evalarg}(a, ent) = \ominus$.

Quiescent A quiescent profile represents an “ideally instinctive” agent evaluating an argument thanks to her S1. More precisely, when receiving an argument, the agent evaluates the argument's conclusion and the speaker. She will accept the argument if she agrees with the conclusion and does not reject the speaker, or vice-versa.

Definition 13 (Quiescent evaluation) Given an argument $a = (s, h, w, (c, f))$, the evaluation of a by a quiescent agent with a cognitive model κ is defined as follows:

$$\text{evalarg}(a, qui) = \begin{cases} \oplus & \text{if } (\text{eval}_1(c, \kappa) = f \text{ and } \text{eval}_1(s, \kappa) \neq \ominus) \text{ or } \\ & (\text{eval}_1(c, \kappa) \neq -f \text{ and } \text{eval}_1(s, \kappa) = \oplus), \\ \ominus & \text{if } (\text{eval}_1(c, \kappa) = -f \text{ and } \text{eval}_1(s, \kappa) \neq \oplus) \text{ or } \\ & (\text{eval}_1(c, \kappa) \neq f \text{ and } \text{eval}_1(s, \kappa) = \ominus), \\ \odot & \text{otherwise} \end{cases}$$

In future work, we plan to take into account the extra sources of persuasion such as the context created by the source of information including trustworthiness and charisma of the source, the contextual mood of the agent, etc.

Example 4 During a long and very technical meeting, when a partner said that “since post harvest separation is highly expensive, which is undesirable, post harvest has to be rejected”, our internship student did not have the cognitive ability to rationally consider this argument. While he would not have agreed with a deeper analysis, he instead relied on his S1, where post harvest separation is associated with something he rejects (see Fig. 1), and therefore accepted the argument.

Engaged An engaged profile represents an “ideally rational” agent evaluating an argument exclusively thanks to her knowledge base. In this work, we propose to define an engaged agent as someone who evaluates an argument wrt its set of warrants that are encoded in a way to capture critical questions (see [32, 4]). An engaged agent has to pass three steps before validating an argument: validity of the warrant (“Am I able to recognize this scheme of thought as a valid one?” translated into “Does it already exists in my personal base of a -rules”)⁶; a syntactic validity of the use of the warrant in the argument (“Is the warrant conform with the premises and conclusions of the argument?” translated in terms of existence of a unification function σ); rational validation of applicability (“Are the premises correct and necessary ?” translated into the use of contextual inference in order to prove them).

Definition 14 (Engaged evaluation) *Given an argument $a = (s, h, w, (c, f))$, with $h = (K_h, A_h)$, the evaluation of a by an engaged agent with a cognitive model $\kappa = (KB, AT, \lambda, i)$ with $KB = (F, \Delta, A, R)$ is defined as follows:*

$$\text{evalarg}(a, \text{eng}) = \begin{cases} \oplus & \text{if } \begin{cases} w \in R \text{ and} \\ \exists \sigma : \mathcal{V}_G \rightarrow \mathcal{V} \text{ s.t. } \sigma(w) = (h \Rightarrow (c, f)) \text{ and} \\ \forall (x \rightsquigarrow y) \in K_h, x \vdash_2 y \text{ and } \neg x \not\vdash_2 y \text{ and } A_h \subseteq A \end{cases} \\ \ominus & \text{if } \begin{cases} w \in R \text{ and} \\ \nexists \sigma : \mathcal{V}_G \rightarrow \mathcal{V} \text{ s.t. } \sigma(w) = (h \Rightarrow (c, f)) \text{ or} \\ \exists (x \rightsquigarrow y) \in K_h, x \not\vdash_2 y \text{ or } \neg x \vdash_2 y \text{ or } A_h \not\subseteq A \end{cases} \\ \odot & \text{otherwise} \end{cases}$$

Example 5 *Several days after the meeting, our internship student thought of the partner’s argument again. Now that he is able to analyze the argument more rationally, he can recognize its type ($w \in R$): his set of warrants R contains two a -rules, $w_1 = (\{a \rightsquigarrow b\}, \{(b, \ominus)\}) \Rightarrow (a, \ominus)$ and $w_2 = (\{a \rightsquigarrow b\}, \{(b, \oplus)\}) \Rightarrow (a, \oplus)$ which encode the schemes associated to arguments from positive or negative consequences (see [32] for a definition of these argumentation schemes). Since $h = (\{phs \rightsquigarrow exp\}, \{(exp, \ominus)\})$ and the conclusion is (phs, \ominus) , the argument is well formed wrt w_1 ; however, w_2 is not applicable. Then, he checks if the premise holds: as seen in Ex. 3, $phs \vdash_2 \neg exp$, and thus $phs \not\vdash_2 exp$. Hence, he rejects the argument.*

3.3 Argument Influence on the Agent’s Mind

Once the argument has been evaluated by an agent, her cognitive model may have to be modified to account for the persuasive impact of the argument. Such modifications can either be the change of a flag value, the addition of a new association or the addition of a new rule. Def. 15 gives the functions representing these modifications.

Definition 15 (Update operations) *Given two cognitive states $\kappa = (KB, AT, \lambda, i)$ with $KB = (F, \Delta, A, R)$ and κ' , two formulae $x, y \in \mathcal{L}$, a set of default rules $D \subseteq \mathcal{L} \times \mathcal{L}$ and a flag $f \in \{\oplus, \ominus, \odot\}$, we define:*

- $\text{noop}(\kappa) = \kappa$
- $\text{setflag}(\kappa, x, f) = \kappa'$ where $\kappa' = ((F, \Delta, A', R), AT', \lambda, i)$ with
 - $L_{AT'} = L_{AT} \cup \{x\}$,
 - $\forall \varphi \in L_{AT} \text{ s.t. } \varphi \neq x, \text{flag}_{\kappa'}(\varphi) = \text{flag}_{\kappa}(\varphi) \text{ and } AT'(\varphi) = AT(\varphi)$,
 - $\text{flag}_{\kappa'}(x) = f \text{ and } A' = A \setminus \{(x, \text{flag}_{\kappa}(x))\} \cup \{(x, f)\} \text{ and } AT'(x) = AT(x)$.
- $\text{push}(\kappa, (x, y)) = \kappa'$ where $\kappa' = (KB', AT', \lambda, i)$ with
 - if $x \notin L_{AT}$ then $AT' = AT \cup \{(x, S_x, \odot)\}$ with $S_x = \{y\}$,
 - else
 - * $\forall \varphi \in L_{AT} \text{ s.t. } \varphi \neq x, \text{flag}_{\kappa'}(\varphi) = \text{flag}_{\kappa}(\varphi) \text{ and } AT'(\varphi) = AT(\varphi)$,
 - * $\text{flag}_{\kappa'}(x) = \text{flag}_{\kappa}(x) \text{ and } AT'(x) = AT(x) \cup \{y\} \text{ with } \text{Top}(x) = y$,
- $\text{addrule}(\kappa, D) = \kappa' \text{ s.t. } \kappa' = (F, \Delta \cup D, A, R), AT, \lambda, i$.

Depending on the profile, the cognitive model will be modified in different ways. These differences aim at representing the fact that the persuasion may be deeper depending on the cognitive involvement of the

⁶Note that we propose to be neutral wrt an argument that uses an unknown warrant.

profile(c, κ)	evalarg($(s, h, w, (c, f))$)		
	\odot	\ominus	\oplus
<i>unc</i>	$\text{push}(\kappa, (c, h))^7$	\times	\times
<i>ent</i>	\times	$\text{push}(\kappa, (c, h))$ $\text{setflag}(\kappa, s, \ominus)$	$\text{push}(\kappa, (c, h))$ $\text{push}(\kappa, (h, c))$ $\text{setflag}(\kappa, s, \oplus)$
<i>qui</i>	$\text{push}(\kappa, (c, h))$	$\text{push}(\kappa, (c, h))$ $\text{setflag}(\kappa, c, -f)$ $\text{setflag}(\kappa, s, \ominus)$	$\text{push}(\kappa, (c, h))$ $\text{push}(\kappa, (h, c))$ $\text{setflag}(\kappa, c, f)$ $\text{setflag}(\kappa, s, \oplus)$
<i>eng</i>	noop	noop	$\text{addrule}(\kappa, K_h)$ $\text{setflag}(\kappa, c, f)$

Table 1: Update of a cognitive state κ .

agent. Table 1 gives the functions to apply to κ in order to update it, according to the possible evaluations of an argument by an agent and her profile. The “ \times ” in the *ent* and *unc* lines corresponds to impossible cases due to, respectively, Postulate 1 and Def. 11.

4 Properties and Postulates

We have not yet been able to experiment in presence of human subjects in order to validate our model, but we have started to explore its rational properties.

4.1 Entailment Properties

Let us examine the properties of S1 and S2-entailment. Due to the construction of \vdash_2 on the basis of contextual entailment, it follows that \vdash_2 is obeying the same properties.

Proposition 1 \vdash_2 obeys the axiom and the five inference postulates of [17]:

- Reflexivity: $a \vdash_2 a$,
- Left logical equivalence: if $\vdash a \leftrightarrow b$ and $a \vdash_2 c$ then $b \vdash_2 c$,
- Right weakening: if $a \vdash b$ and $c \vdash_2 a$ then $c \vdash_2 b$,
- Cut: if $a \wedge b \vdash_2 c$ and $a \vdash_2 b$ then $a \vdash_2 c$,
- Cautious monotony: if $a \vdash_2 b$ and $a \vdash_2 c$ then $a \wedge b \vdash_2 c$,
- Or: if $a \vdash_2 c$ and $b \vdash_2 c$ then $a \vee b \vdash_2 c$.

It is not the same for \vdash_1 , since it may be sensitive to the syntax, i.e., nothing prevents to have a different stack for two equivalent formulae.

Proposition 2

- \vdash_1 obeys Reflexivity only for the formulae that admit finite reflection paths
- \vdash_1 obeys Left logical equivalence only if AT is syntax dependent i.e. $\varphi \leftrightarrow \psi$ iff $AT(\varphi) = AT(\psi)$,
- \vdash_1 does not obey Right weakening
- Transitivity holds, namely, $a \vdash_1 b$ and $b \vdash_1 c$ implies $a \vdash_1 c$
- Cut, Cautious monotony and Or: do not necessarily hold.

Proof: Reflexivity: if $\exists R_a$ s.t. $\text{flag}(R_a) \neq \odot$ then $a \in R_a$ hence $a \vdash_1 a$ otherwise it is not the case. Right weakening: since b can be deducible logically from a but not in $AT(a)$. Transitivity: it means that $b \in R_a$ and $c \in R_b$, hence if $\text{flag}(b) = \odot$ then $c \in R_a$ else $c = b$ hence $c \in R_a$ as well. Cut, Cautious Monotony and Or: it is due to the independence of associations wrt logic (hence “logical and” is not necessarily compatible with associations), \square

⁷An argument is never evaluated by an unconcerned agent. However, we represent the fact that, like enthusiastic and quiescent agents, she is unconsciously influenced by what she hears.

4.2 Incorporation Property

Let us notice that after receiving an argument, the knowledge of an agent can only increase: more precisely, among the formulae that were already present, the number of flags that are not empty decreases (however some new formula may be added with an empty flag) and the number of associations grows. Moreover some rules can also be added in the case of an engaged profile.

Proposition 3 *Let $\kappa = ((F, \Delta, A, R), AT, \lambda, i)$, $\kappa' = ((F', \Delta', A', R'), AT', \lambda', i')$ such that κ' is the cognitive model obtained from κ after the utterance of an argument. It holds that $L_{AT} \subseteq L_{AT'}$, $\forall \varphi \in L_{AT}, AT(\varphi) \subseteq AT'(\varphi)$, and $F = F', \Delta \subseteq \Delta', R = R', \lambda = \lambda'$ and $i = i'$.*

Note that the flag values are non-monotonic since a formula can obtain either an accepted, rejected or empty flag depending on the engagement profile.

4.3 Public Opinion Axioms

According to [34], the model of how information is transformed in public opinion follows four axioms mentioned below. Our proposal satisfies these axioms:

Reception axiom: *The greater the level a person's level of cognitive engagement with an issue the more likely he / she will be exposed to and comprehend political messages concerning that issue.* It holds since an unconcerned agent does not evaluate the argument, an enthusiastic agent takes it into account if she agrees with the conclusion, a quiescent agent evaluates it with S1-reasoning and an engaged agent evaluates it with S2-entailment. Hence, the more engaged an agent is, the more information she takes into account (in the following order: unconcerned, enthusiastic, quiescent, engaged).

Resistance axiom: *People tend to resist arguments that are inconsistent with their political predispositions but they do so only to the extent that they possess the contextual information necessary to perceive a relationship between the message and their predispositions.* Unconcerned, enthusiastic and engaged agents may resist an argument since they are not influenced by its flag. A quiescent agent resists arguments that are against her opinion or uttered by a source she rejects (see Definition 13).

Accessibility axiom: *The more recently a consideration has been called to mind, or thought about, the less time it takes to retrieve that consideration or related considerations from memory and bring them to the top of the head for use.* This axiom is satisfied concerning the association table AT since every kind of profile add the new piece of information at the top of the stack (see Table 1).

Response axiom: *Individuals answer survey questions by averaging across the considerations that are immediately salient or accessible to them.* It holds for quiescent and enthusiastic: a quiescent agent evaluates a formula by considering the most immediately accessible information and an enthusiastic agent evaluates only the immediate value of a formula. However, it does not hold for unconcerned and engaged agents: one does not evaluate the formula, and the other evaluates the formula with her knowledge base.

5 Conclusion

This paper is a first proposal of a formalization of dual process theory and its link with human persuasion. Based on the ELM model of persuasion, we define four profiles evaluating an argument in different ways. One of the profiles aims at reasoning thanks to an association table, and another is based on a logical inference mechanism named contextual entailment. This mechanism is a possible implementation of S2 and can be changed without jeopardizing the cognitive model. Moreover, each profile integrates the contents of the received argument differently. Accordingly to public opinion axioms, the more cognition was involved in its evaluation, the more persuasive content will take root in the mind of the agent.

5.0.1 Related Work

Dual process theories have already been implemented for problem solving. Namely, [14] with an extension of the CLARION architecture that relies on two modules: a bottom-level (resp. top-level) module handling implicit knowledge (resp. explicit knowledge), which recall the S1 and S2 systems but is not based on formal logic. [27] proposes a general intelligence cognitive architecture composed of a long-term memory

independent of specific tasks and a capacity-limited working memory. The S1 and S2 systems allow them to distinguish between perception and imagination and are represented thanks to two binary relations on the element of the long-term memory and two propagation processes. Some works, similarly to ours but not in a logical framework, aim at explaining purely human processes. For instance, [18] studies the emergence of emotions thanks to a three-levels cognitive architecture: S1 (the *reactive* level) and S2, subdivided into the *algorithmic* level and the *reflective* level. The first one is responsible for fast and instinctive behaviours, the second one is used for cognitive control and the last one handles rational behavior.

[16] is a different approach for persuasion, since the NAG program is able to analyze and generate arguments with the aim of persuading a human user. In order to do so, NAG comprises two different models, a normative model that is able to judge the correctness of an argument (in terms of links between the premises and the conclusion), and a user model, that is able to evaluate the persuasion capability of an argument on the user. Hence, NAG is interestingly able to analyze an argument given by the user and to try to generate a counterargument which is at the same time correct and specifically designed to be effective on the user. Since NAG has to persuade a human user, it requires a representation of her cognitive profile, in particular her reasoning errors such as cognitive biases. Major differences exist between our approach and NAG. First, NAG is intended to interact with users, and as such it is human-machine oriented. Then, the model does not rely on a logical dual process but is based on a Bayesian network; cognitive biases are thus taken into account by the modification of probability degrees while, in our framework, biases are due to faulty appreciations, warrants or beliefs. Finally, the authors do not use argumentation schemes (encoded in our warrants base R) and thus do not have a clear definition of argument and ways to evaluate them.

5.0.2 Perspectives

Since this work is a first attempt to formalize a two-process cognitive model and its links with argument evaluation, numerous perspectives can be envisaged. Namely, a refined definition of the weights associated to the reflection paths could help to account for the various heuristics Kahneman and Tversky described. Moreover, we would like to investigate a way to compute the cognitive availability of an agent in order to determine her ability to engage in the argument evaluation. Such study would benefit from the definition of a profile corresponding to a continuum between the quiescent and the engaged profiles. Such a profile might be more adapted to represent human reasoning with its bounded rationality, and may help to better capture the difference of persuasion strategies that can be used according to the proximity to S1 or S2-reasoning. Moreover, a more realistic model should take into account social influence [7] between agents when exchanging arguments. In the same way, the study of rhetorical mechanisms could improve the evaluation of argument with another dimension. Finally, the public opinion axioms of [34] show that results from psychology studies can be used to guide our research, and as such it would be useful to validate our proposal by an empirical study with human beings.

References

- [1] L. Amgoud, N. Maudet, and S. Parsons. An argumentation-based semantics for agent communication languages. In *ECAI'02*, pages 38–42. IOS Press, 2002.
- [2] C. G. Beevers. Cognitive vulnerability to depression: A dual process model. *Clinical Psychology Review*, 25(7):975 – 1002, 2005.
- [3] S. Benferhat and F. Dupin de Saint Cyr - Bannay. Contextual handling of conditional knowledge. In *Proc. of IPMU'96*, Granada, Spain, July 1996.
- [4] J. A. Blair and R. H. Johnson. Informal logic: An overview. *Informal Logic*, 20(2), 2000.
- [5] R. Burrows, H. Johnson, and P. Johnson. Developing an online social media system to influence pro-environmental behaviour based on user values. In *ICPT*, 2014.
- [6] S. Chaiken. The heuristic model of persuasion. *Social influence: The Ontario Symposium*, 5:3–37, 1987.
- [7] R. Cialdini. *Influence: Science and Practice*. Allyn and Bacon, 2001.

- [8] C. S. Clements. Perception and Persuasion in Legal Argumentation: Using Informal Fallacies and Cognitive Biases to Win the War of Words. *BYU Law Review*, 319, March 2013.
- [9] P. Croskerry, G. Singhal, and S. Mamede. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Quality & Safety*, 22(Suppl 2):58–64, 2013.
- [10] S. Epstein. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49(8):709, 1994.
- [11] J. S. B. T. Evans and J. Curtis-Holmes. Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11(4):382–389, 2005.
- [12] A. Forget, S. Chiasson, P. van Oorschot, and R. Biddle. Persuasion for stronger passwords: Motivation and pilot study. In *Persuasive Technology*, volume 5033 of *LNCIS*, pages 140–150. Springer Berlin Heidelberg, 2008.
- [13] C. Hamblin. *Fallacies*. University paperback. Methuen, 1970.
- [14] S. Hélie and R. Sun. Incubation, insight, and creative problem solving: a unified theory and a connectionist model. *Psychological Review*, 117(3):994–1024, 2010.
- [15] J. Hornikx and U. Hahn. Reasoning and argumentation: Towards an integrated psychology of argumentation. *Thinking & Reasoning*, 18(3):225–243, 2012.
- [16] K. B. Korb, R. Mcconachy, and I. Zukerman. A cognitive model of argumentation. In *Proc. of the Nineteenth Annual Conference of the Cognitive Science Society*, pages 400–405, 1997.
- [17] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.
- [18] O. Larue, P. Poirier, and R. Nkambou. Emotional emergence in a symbolic dynamical architecture. In *BICA 2012*, volume 196 of *AISC*, pages 199–204. Springer, 2013.
- [19] T. Lehto and H. Oinas-Kukkonen. Explaining and predicting perceived effectiveness and use continuance intention of a behaviour change support system for weight loss. *Behaviour & Information Technology*, 34(2):176–189, 2015.
- [20] J. Mackenzie. Four dialogue systems. *Studia Logica*, 49(4):567–583, 1990.
- [21] H. Oinas-Kukkonen. A foundation for the study of behavior change support systems. *Personal and Ubiquitous Computing*, 17(6):1223–1235, 2013.
- [22] C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, June 1969.
- [23] R. Petty and J. Cacioppo. The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, 19(C):123–205, 1986.
- [24] H. Prakken. Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 21(2):163–188, June 2006.
- [25] I. Rahwan and L. Amgoud. An argumentation based approach for practical reasoning. In *Proc. of Int. J. conf. on Autonomous Agents and MultiAgent Systems*, pages 347–354, 2006.
- [26] S. A. Sloman. The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1):3, 1996.
- [27] C. Strannegard, R. von Haugwitz, J. Wessberg, and C. Balkenius. A cognitive architecture based on dual process theory. In *AGI*, volume 7999 of *LNCIS*, pages 140–149. Springer, 2013.

- [28] D. Touretzky. Implicit ordering of defaults in inheritance systems. In *Proc. of AAAI'84*, University of Texas at Austin, 1984.
- [29] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.
- [30] D. van Knippenberg. Social identity and persuasion: Reconsidering the role of group membership. *Social identity and social cognition*, XVII:315–331, 1999.
- [31] D. Walton. *Logical Dialogue: Games and Fallacies*. University Press of America, 1984.
- [32] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, Cambridge, 2008.
- [33] W. Wood. Attitude change: Persuasion and social influence. *Annual review of psychology*, 51(1):539–570, 2000.
- [34] J. Zaller. *The Nature and Origins of Mass Opinion*. Cambridge Studies in Political Psychology Series. Cambridge University Press, 1992.